The Application of Gaze Heat Maps in Impression Evaluation : Case Study of Chinese College Students Evaluating Historical Souvenir Stores in Japan

LIU, Chenyang
Graduate School of Integrated Frontier Sciences, Kyushu University

SOGABE, Haruka
Kyushu University

https://hdl.handle.net/2324/7405101

出版情報：International Journal of Affective Engineering. 19 (1), pp.67-77, 2020. 日本感性工学会
バージョン：

International Journal of Affective Engineering Vol.19 No.1 pp.67-77 (2020)
doi: 10.5057/ijae.TJSKE-D-19-00019

ORIGINAL ARTICLE

The Application of Gaze Heat Maps in Impression Evaluation

– Case Study of Chinese College Students Evaluating Historical Souvenir Stores in Japan –

Chenyang LIU* and Haruka SOGABE**

** Graduate School of Integrated Frontier Sciences, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Fukuoka 815-8540, Japan*
*** Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Fukuoka 815-8540, Japan*

Abstract: Eye trackers have gradually been applied in research including in the Kansei Engineering field. Furthermore the semantic differential (SD) method can obtain users' Kansei from different aspects through the semantics of different words. As for facade, because it contains too many components, the components' effect on Kansei is difficult to analyze using only evaluation scores. In this paper, an eye tracker was utilized in the facade evaluation process to analyze whether evaluation words would have an effect on gaze. Subjects were divided into four groups, three of which were given different evaluation words, and the fourth group was not given any evaluation word. The differences in heat maps were quantified and compared. While using evaluation words in the Kansei evaluation experience, the heat maps were more similar than those without using any evaluation word, indicating that the logic of the evaluation task, the Kansei evaluation words, and individual preference all had an effect on gaze position.

Keywords: *Semantic differential method, Eye tracking, Heat map*

1. INTRODUCTION

The semantic differential (SD) method is a rating scale designed to measure the connotation of objects; it is often deployed in Kansei Engineering (KE) research, and has proven to be effective [1]. However, the method has its own limitations; for example, a single evaluator has to evaluate several design works, which may place a heavy burden on the evaluator; furthermore, on-site evaluation is difficult to implement for facade or landscape designs. Therefore, it is necessary to improve the evaluation efficiency of the SD method utilized in Kansei evaluation.

Although extensive research has been conducted on facade design evaluation based on the SD method, the majority has focused on comparing different facades to identify the components that affect the Kansei of the facades. However, facades have many components, and similar components may have different effects on different facades.

Eye trackers can be utilized in various fields such as KE and human–computer interaction (HCI) [2-4], and they can provide deeper insight into user's observation process, making it possible to obtain more information from a single evaluation process.

Many indices exist for eye movement analysis. Using heat mapping, the user's range and area of focus can be accurately obtained. Furthermore, heat mapping can intuitively visualize subjects' attention and the time spent observing each part of a photograph based on the gaze results. Although the data inherent in the heat map are abundant, it is difficult to analyze them quantitatively. However, it is achievable through programming. In this study, the Kansei evaluation of facades were conducted using heat mapping.

This study utilized an eye tracker for the Kansei evaluation of store facades using screen-based facade evaluation and discusses the analysis method of data acquired using heat map obtained from the eye tracker used to monitor the SD evaluation.

The subjects of this study were Chinese students assigned the task of evaluating souvenir stores around Japan's historical tourist areas. Recently, there has been an annual increase in the number of Chinese tourists visiting Japan; these tourists are often in their 20s, a group that likes to share travel photos through social networks. This study aims to investigate the possibility of using an eye tracker for the Kansei evaluation of the facade design by analyzing the influence of different evaluation words on gaze.

2. THE RELATIONSHIP BETWEEN EVALUATION WORDS AND GAZE POSITION

2.1 Method

Research on HCI focuses on the interfaces between users and computers. To deploy eye trackers in HCI research, subjects are usually assigned certain tasks. Gaze, as recorded by an eye tracker, is analyzed to identify common points, and is thereafter compared with tasks [4, 5]. However, because the tasks assigned to subjects in these

Received: 2019.04.29 / Accepted: 2019.10.26
J-STAGE Advance Published Date: 2019.12.26

experiments often have a clear focus, the gaze patterns of the different subjects are usually very similar. This is attributable to the similarity in the logic of the tasks; therefore, regardless of individual cognition, the gaze patterns of different subjects executing the same task will have some similarities [4]. However, some studies have observed that subjects' gaze patterns may vary due to individual cognition [6]. The similarity in the gaze patterns of different subjects is attributable to what we subsequently refer to as "universal thinking". Universal thinking may be generated due to the similarity in the logic of tasks assigned to different subjects. The difference in the gaze patterns of different subjects assigned similar evaluation tasks is attributable to what we subsequently refer to as "individual thinking", arising from individual preference and cognition.

In the SD method, first, different evaluation words are provided to the subjects, who observe the evaluation objects (physical objects or photographs) and rate them [4]. Therefore, it can be conjectured that logic plays a role when a subject is assigned an evaluation task. The gaze positions of the subjects are also influenced by universal thinking and individual thinking. However, tasks based on the SD evaluation differ from tasks in HCI field experiments. The SD method investigates the different sensations arising from different evaluation words given to the subjects, which have no clear link to the evaluated objects. Although we know that there are similarities in the Kansei of different users, it is yet unknown whether this similarity may be gleaned from gaze positions.

During the evaluation process of subjects using the SD method, an eye tracker is used to record the observation process, and the recorded data are analyzed to investigate the influence of unique evaluation words assigned to the different subjects before the experiment on their gaze patterns.

To identify the relationship between the evaluation words and gaze position during evaluation, Kansei evaluation based on the SD method was conducted using random facade photographs, as shown in Figure 1, which were taken on a sunny day with the approval of a store in the Kiyomizu area of Kyoto. The gaze patterns of the subjects,

Figure 1: Facade photograph

as recorded by the eye tracker, were utilized throughout the experiment. The experiment was ethically examined and approved by Kyushu University. The approval number was 310.

Tobii Pro Nano (from Tobii pro AB, 2018), with a sampling rate of 60 Hz, head movable range of 35*30 cm, and a 19 inch 16:9 screen was used to collect eye movements. The subjects were 45 cm away from the screen, following a visual acuity test established that they could clearly see the pictures and words on the screen from this distance.

(1) Error Measurement Experiment

Errors were measured before the actual experiment.

The measurement method: A blue circle with a diameter of 5 pixels appeared at random positions on the screen and blinked for 3 s. Next, a white circle of the same size appeared at the same spot for 1 s. The subjects were instructed to devote unwavering focus on the white circle. After 1 s, the white circle disappeared, and the screen remained blank for 5 s. The process commencing from the appearance of the blue circle to the screen blankness counted as one set of the experiment.

The eye tracker consistently recorded the gaze coordinate when the white circle appeared. The program used in the experiment recorded the coordinate of the white circle. Because the sampling rate of the eye tracker was 60, for one set of the experiment, 60 gaze coordinates and one circular coordinate could be recorded. Twenty subjects participated in this experiment, and each conducted 60 sets of the experiment.

The deviation distance between the coordinate of the white circle and that of the gaze was calculated. According to the results, the distance deviations in the horizontal and vertical directions coincided with the normal distribution (horizontal deviation $\mu \approx 0.13$, $\sigma \approx 11.33$; vertical deviation $\mu \approx 0.18$, $\sigma \approx 9.87$).

When the gaze coordinates acquired using the eye tracker was (x_0, y_0), the observation probability P of the subjects' actual observation coordinate (x_1, y_1) was calculated as:

$$\Delta x = x_1 - x_0 \qquad \Delta y = y_1 - y_0$$

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi} \times \sigma} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P = f(\Delta x; 0.13, 11.33) \times f(\Delta y; 0.18, 9.87)$$

A large amount of calculation was necessary (screen resolution: 1920*1080, 2,073,600 coordinates in total), to simplify calculation, when the P of a coordinate was less than 0.00007, it was discarded. The sum value of the discarded P was approximately 0.05.

(2) Evaluation Experiment

During the experiment, the subjects maintained a sitting posture. The experiment was conducted in a quiet indoor environment with sufficient lighting. The subjects were instructed beforehand to avoid large body and head movements. If a subject's body or head was found to move beyond the eye tracker's measurement range, the results were discarded. The 48 subjects (24 males and 24 females) were college students majoring in design who had never visited Japan; they were between the ages of 20–23 years.

To confirm whether the different evaluation words affected the gaze of the subjects, the subjects were divided into four groups of 12 (six males and six females); three groups were told before the experiment to observe the store's "Beautiful-Ugly", "Close-Distant" and "Classical-Modern" features, respectively, and subsequently to fill the evaluation sheet following their observation. No evaluation word was provided to the fourth group beforehand. The experiment began with calibration conducted using the official software Tobii Pro (Tobii Pro Eye Tracker Manager, Ver.1.12.2) after the provision of the evaluation words. The calibration took approximately 90 s per person. Afterward, the store facade photograph was automatically displayed for 120 s. The software used for observation was a program written in Python based on the official Tobii Pro SDK (Ver.1.6) that could commence data collection while the picture was being shown and stop data collection and dismiss the picture after 120 s.

After the exercise, an interview was conducted in which the subjects were to talk about the impressive objects in terms of the evaluation words (if no evaluation words had been provided, they were simply asked about their overall impression); the interview lasted approximately 3 minutes; then, they filled the evaluation sheet. For the evaluation words, three words commonly used in facade evaluation research were selected from the research on facades based on the SD method from the *Transactions of Japan Society of Kansei Engineering* and the *International Journal of Affective Engineering*.

(3) Heat Map Generation and Analysis Method

According to the error measurement experiment, when the eye tracker obtained a coordinate, the observation probabilities of this coordinate and all the coordinates in an ellipse centered on this coordinate were superposed. The range of the ellipse is shown in Figure 2. The sum of the *P* of all the coordinates in the ellipse was approximately 0.95. After all the single-coordinate-heat maps were computed numerically, the P of each pixel in each single-coordinate heat map was superposed to generate a heat map. The superimposed P values were subsequently referred to as the dynamic values. The pictures are downscaled on the screen, because the aspect ratio of the photographs used in the experiment differed from that of the screen. Therefore, the dynamic values outside the pictures were deleted. Finally, different colors and transparency were used to represent different dynamic values to complete the heat map. The heat maps were generated and calculated using a program written by Python.

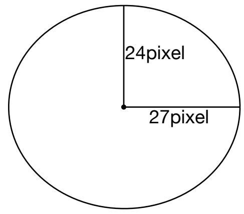


Figure 2: Range of the ellipse

Furthermore, because the data were arranged in chronological order, time was a significant variable in the analysis of the heat map. To understand the evaluation process of the subjects, 60 heat maps numbered 1–60 were generated from the data recorded in a single experiment. The heat map, numbered n, contains all the data for 0–2n s. This series of heat maps is tagged the chronological heat map.

Although we can clearly observe the gaze results of a single subject from the heat map, a comparative analysis is challenging. To analyze the heat map quantitatively, the dynamic values corresponding to each coordinate were utilized. The differences in the heat maps of the different subjects during the same time period were calculated using the Euclidean distance method, and were compared, to understand the extent to which they differed. At each time period, the Euclidean distance D_{AB} between heat maps *A* and *B* is calculated as:

$$D_{AB} = \sqrt{\sum_{(x,y)} (A_{(x,y)} - B_{(x,y)})^2}$$

$$0 < x \leqslant 1920 \quad 0 < y \leqslant 1080$$

where (x, y) is the coordinate in the picture, and $A(x, y)$ represents the dynamic value of heat map A in coordinate (x, y). Based on this method, the Euclidean distance (hereinafter referred to as the differential distance) among the heat maps of all the subjects in the same time period could be calculated. The distance between the two heat maps is directly proportional to the difference between them.

(4) Testing Method

To investigate whether the evaluation words influenced the gaze position, the calculated differential distance of the heat maps in the same group and different groups were used for a statistical test. One-way ANOVA was used, treating the group as an independent variable and the differential distance as a dependent variable. If there was a significant difference between the groups, post hoc test with LSD method was conducted.

2.2 Result

A total of 48 sets of eye movement data were obtained. Each set contained 60*120 coordinates arranged in chronological order. Forty-eight sets of interviews were conducted. A total of 2,880 (12*4*60) heat maps were generated in the experiments.

The gaze of each subject at the end of the 120 s experiment was converted into a heat map, as shown in Figure 3, which, due to layout limitations, shows only the heat maps of the group that were supplied with the word "Beautiful-Ugly". Differences were observed in the heat maps.

(1) The Differential Distance Between Subjects in the Same Group

The differential distance between the experimental groups and the control group were calculated and compared. The test results are shown in Table 1, and the post hoc test results are shown in Table 2.

Mostly, there were significant differences between the experimental group and the control group. Significant differences were noted between the group with the evaluation word "Beautiful-Ugly" and the control group across all the time periods, besides 4–8 s, 18 s, 22–24 s, 30–36 s, and the time periods after 102 s. A significant difference was noted between the group with the evaluation word "Close-Distant" and the control group at the end of 2 s, 10–14 s, 26–28 s, 36 s, and 40–66 s. There was no significant difference between these two groups after 66 s. Significant differences were noted between the group with the evaluation word "Classical-Modern" and the control group at 10–14 s and 26–90 s. Overall, significant differences could be noted at the end of 40–66 s in all three groups.

Figure 3: Generated heat map with evaluation word "Beautiful-Ugly"

Table 1: Differential distance comparison between subjects in the same group

Time (s)	F-value	Time (s)	F-value	Time (s)	F-value	Time (s)	F-value
	df = (3, 260)		df = (3, 260)		df = (3, 260)		df = (3, 260)
2	5.16**	32	7.52**	62	5.78**	92	9.48**
4	2.74*	34	8.40**	64	6.28**	94	7.08**
6	0.97	36	10.89**	66	6.87**	96	7.14**
8	0.46	38	12.58**	68	9.16**	98	7.27**
10	7.32**	40	12.80**	70	10.08**	100	7.39**
12	10.84**	42	14.52**	72	11.11**	102	6.69**
14	13.81**	44	18.38**	74	11.22**	104	6.37**
16	9.34**	46	19.77**	76	11.96**	106	5.36**
18	2.79*	48	20.91**	78	12.81**	108	5.84**
20	4.46**	50	19.81**	80	13.71**	110	6.51**
22	2.34	52	16.72**	82	15.47**	112	6.25**
24	2.84*	54	13.69**	84	16.80**	114	5.62**
26	6.52**	56	11.12**	86	18.41**	116	5.12**
28	5.90**	58	8.40**	88	20.00**	118	5.15**
30	7.57**	60	6.07**	90	15.33**	120	5.73**

*p<.05, **p<.01

Table 2: Post-hoc test results

Time (s)	Mean (SD) of BU	Mean (SD) of CD	Mean (SD) of CM	Mean (SD) of N	Time (s)	Mean (SD) of BU	Mean (SD) of CD	Mean (SD) of CM	Mean (SD) of N
2	1.2(0.2)*	1.2(0.2)*	1.2(0.4)	1.0(0.2)	62	11.4(0.9)*	11.7(1.4)*	11.6(1.7)*	12.7(1.6)
4	1.9(0.2)	2.0(0.3)	1.9(0.3)	1.8(0.3)	64	11.7(1.0)*	12.0(1.4)*	11.9(1.7)*	13.1(1.7)
6					66	12.1(1.1)*	12.5(1.4)*	12.1(1.6)*	13.5(1.9)
8					68	12.3(1.0)*	13.3(1.5)	12.3(1.6)*	14.0(2.2)
10	3.3(0.4)*	3.3(0.3)*	3.3(0.3)*	3.7(0.5)	70	12.5(1.1)*	13.7(1.6)	12.6(1.6)*	14.3(2.3)
12	3.6(0.4)*	3.7(0.3)*	3.8(0.3)*	4.1(0.6)	72	12.7(1.1)*	14.1(1.7)	12.9(1.6)*	14.7(2.4)
14	3.9(0.4)*	4.1(0.2)*	4.2(0.4)*	4.6(0.6)	74	12.9(1.1)*	14.5(1.8)	13.2(1.7)*	14.9(2.5)
16	4.4(0.4)*	4.5(0.3)	4.8(0.6)	4.9(0.6)	76	13.2(1.2)*	15.0(1.8)	13.5(1.8)*	15.3(2.6)
18	4.9(0.5)	5.0(0.5)	5.1(0.6)	5.2(0.6)	78	13.4(1.3)*	15.5(1.9)	13.8(1.9)*	15.6(2.6)
20	5.1(0.5)*	5.3(0.5)	5.5(0.7)	5.5(0.6)	80	13.7(1.3)*	16.0(2.0)	14.1(1.9)*	16.0(2.7)
22					82	13.9(1.4)*	16.4(2.0)	14.3(2.0)*	16.3(2.8)
24	5.9(0.5)	5.9(0.6)	6.0(0.8)	6.3(0.8)	84	14.1(1.4)*	16.8(2.0)	14.7(1.9)*	16.6(2.8)
26	6.1(0.5)*	6.2(0.6)*	6.2(0.7)*	6.7(0.8)	86	14.2(1.3)*	17.2(2.1)	15.0(1.9)*	16.8(2.8)
28	6.5(0.7)*	6.6(0.7)*	6.5(0.7)*	7.1(0.8)	88	14.4(1.4)*	17.7(2.1)	15.4(1.9)*	17.2(2.8)
30	6.9(0.7)	7.1(0.7)	6.7(0.8)*	7.6(1.0)	90	15.0(1.6)*	18.0(2.1)	16.0(2.0)*	17.5(2.8)
32	7.3(0.8)	7.4(0.7)	6.9(0.8)*	7.9(1.2)	92	15.8(1.9)*	18.2(2.1)	16.5(2.1)	17.8(2.8)
34	7.7(0.9)	7.8(0.8)	7.2(0.8)*	8.3(1.2)	94	16.4(1.9)*	18.5(2.1)	17.0(2.2)	18.2(2.8)
36	8.1(1.0)	8.2(0.9)*	7.5(0.9)*	8.7(1.3)	96	16.8(2.1)*	18.9(2.3)	17.2(2.2)	18.4(2.8)
38	8.4(0.9)*	8.5(1.0)	7.7(0.9)*	9.1(1.3)	98	17.1(2.2)*	19.3(2.4)	17.6(2.2)	18.7(2.9)
40	8.6(0.9)*	8.7(1.0)*	7.9(1.0)*	9.5(1.4)	100	17.3(2.2)*	19.8(2.8)	18.0(2.3)	19.1(3.1)
42	8.8(0.9)*	9.0(1.0)*	8.2(1.0)*	9.9(1.5)	102	17.6(2.2)*	20.2(3.1)	18.4(2.3)	19.3(3.2)
44	9.1(0.9)*	9.2(1.1)*	8.4(0.9)*	10.3(1.6)	104	18.1(2.4)	20.6(3.4)	18.6(2.3)	19.5(3.2)
46	9.3(0.9)*	9.5(1.2)*	8.6(1.0)*	10.6(1.5)	106	18.4(2.5)	20.9(3.5)	19.1(2.3)	19.7(3.2)
48	9.5(0.9)*	9.8(1.2)*	8.8(1.1)*	10.9(1.5)	108	18.5(2.5)	21.3(3.6)	19.6(2.4)	20.0(3.3)
50	9.8(0.9)*	10.1(1.2)*	9.1(1.1)*	11.1(1.5)	110	18.7(2.5)	21.8(3.6)	20.0(2.5)	20.5(3.5)
52	10.0(0.9)*	10.4(1.3)*	9.4(1.2)*	11.4(1.5)	112	19.1(2.6)	22.2(3.7)	20.4(2.7)	20.9(3.6)
54	10.2(0.9)*	10.6(1.3)*	9.9(1.3)*	11.6(1.6)	114	19.5(2.5)	22.6(4.0)	21.0(3.0)	21.3(3.6)
56	10.4(0.8)*	10.9(1.2)*	10.3(1.4)*	11.8(1.6)	116	19.9(2.5)	22.9(4.0)	21.3(3.1)	21.6(3.7)
58	10.7(0.9)*	11.2(1.3)*	10.8(1.4)*	12.1(1.6)	118	20.1(2.5)	23.1(3.9)	21.7(3.0)	21.9(3.7)
60	11.0(0.8)*	11.5(1.3)*	11.3(1.7)*	12.4(1.6)	120	20.3(2.5)	23.4(3.9)	22.0(3.0)	22.1(3.7)

BU: Differential distance between subjects in group with word "Beautiful-Ugly"
CD: Differential distance between subjects in group with word "Close-Distant"
CM: Differential distance between subjects in group with word "Classical-Modern"
N: Differential distance between subjects in group without evaluation word
*Significant difference between this group and N

The mean of the differential distance in each group increased consistently during evaluation. At 40–66 s, the most marked differences could be observed in all the experimental groups; the mean differential distance of the control group was the greatest. After approximately 80 s, the mean differential distance of the control group was surpassed by that of the "Distant-Close" group.

From this result, it can be deduced that the use of evaluation words affected the subject's gaze position at particular periods of time that varied for all the evaluation words.

(2) The Differential Distance Between Subjects in Different Groups

The differential distance between the subjects in the three experimental groups were calculated. The result of the statistical test for the distance between the subjects in the experimental groups and the distance between the subjects in the control group are shown in Table 3, and the post-hoc test results are shown in Table 4.

The differential distance of subjects in the experimental groups and those in the control group exhibited marked differences at particular periods of time. The means of the differential distances are shown in Table 4. In the time period when there was marked difference, the mean of the differential distance from the different experimental groups was lower than that from the control group, except at the end of 2 s.

Table 3: Differential distance comparison between subjects in the different groups

Time (s)	F-value	Time (s)	F-value	Time (s)	F-value	Time (s)	F-value
df1 = 3 df2 = 260		df1 = 3 df2 = 260		df1 = 3 df2 = 260		df1 = 3 df2 = 260	
2	3.60**	32	6.27**	62	7.92**	92	4.72**
4	1.92	34	6.26**	64	8.13**	94	3.55**
6	0.55	36	7.54**	66	8.38**	96	3.31**
8	0.26	38	8.41**	68	9.40**	98	3.26**
10	5.92**	40	9.06**	70	9.78**	100	3.23**
12	8.67**	42	10.83**	72	9.63**	102	2.88**
14	11.32**	44	14.33**	74	8.98**	104	2.50*
16	6.96**	46	15.73**	76	8.99**	106	2.07
18	2.43*	48	17.13**	78	9.25**	108	2.32*
20	3.30**	50	16.34**	80	9.22**	110	2.61*
22	2.25*	52	14.90**	82	10.09**	112	2.43*
24	3.14**	54	13.87**	84	10.63**	114	2.20*
26	7.05**	56	12.50**	86	10.68**	116	2.00
28	6.06**	58	10.49**	88	10.87**	118	2.11
30	7.12**	60	8.24**	90	8.06**	120	2.37*

*p<.05, **p<.01

Table 4: Post hoc test results

Time (s)	Mean (SD) of BU-CD	Mean (SD) of CM-BU	Mean (SD) of CD-CM	Mean (SD) of N	Time (s)	Mean (SD) of BU-CD	Mean (SD) of CM-BU	Mean (SD) of CD-CM	Mean (SD) of N
2	1.2(0.2)	1.2(0.3)	1.2(0.3)	1.0(0.2)	62	11.4(1.2)*	11.5(1.4)*	11.6(1.4)*	12.7(1.6)
4					64	11.7(1.2)*	11.8(1.4)*	11.8(1.4)*	13.1(1.7)
6					66	12.1(1.3)*	12.1(1.5)*	12.2(1.4)*	13.5(1.9)
8					68	12.6(1.3)*	12.3(1.5)*	12.7(1.4)*	14.0(2.2)
10	3.3(0.4)*	3.3(0.4)	3.3(0.3)	3.7(0.5)	70	12.9(1.4)*	12.5(1.5)*	13.0(1.5)*	14.3(2.3)
12	3.6(0.4)*	3.7(0.4)*	3.7(0.3)*	4.1(0.6)	72	13.3(1.5)*	12.8(1.5)*	13.4(1.6)*	14.7(2.4)
14	4.0(0.3)*	4.1(0.4)*	4.2(0.3)*	4.6(0.6)	74	13.6(1.5)*	13.1(1.5)*	13.7(1.6)*	14.9(2.5)
16	4.4(0.4)	4.5(0.6)	4.6(0.5)	4.9(0.6)	76	14.0(1.6)*	13.4(1.6)*	14.1(1.7)	15.3(2.6)
18	4.9(0.5)	4.9(0.6)	5.0(0.5)	5.2(0.6)	78	14.3(1.7)*	13.6(1.7)*	14.5(1.8)	15.6(2.6)
20	5.1(0.5)	5.2(0.6)	5.3(0.6)	5.5(0.6)	80	14.7(1.8)	13.9(1.7)*	14.9(1.8)	16.0(2.7)
22	5.6(0.5)*	5.6(0.6)	5.7(0.6)	5.9(0.7)	82	15.0(1.8)	14.1(1.7)*	15.2(1.9)	16.3(2.8)
24	5.8(0.5)	5.8(0.7)*	5.9(0.7)	6.3(0.8)	84	15.4(1.8)	14.4(1.7)*	15.6(1.8)	16.6(2.8)
26	6.1(0.6)*	6.1(0.7)*	6.2(0.7)*	6.7(0.8)	86	15.7(1.9)	14.7(1.7)*	16.0(1.9)	16.8(2.8)
28	6.5(0.7)*	6.4(0.7)*	6.5(0.7)*	7.1(0.8)	88	16.0(1.9)	15.0(1.8)*	16.4(1.9)	17.2(2.8)
30	6.9(0.7)	6.7(0.8)*	6.9(0.7)*	7.6(1.0)	90	16.4(2.0)	15.6(2.0)*	16.8(2.0)	17.5(2.8)
32	7.3(0.8)	7.0(0.9)*	7.1(0.7)*	7.9(1.2)	92	16.9(2.0)	16.3(2.2)*	17.2(2.1)	17.8(2.8)
34	7.7(0.9)	7.4(0.9)*	7.5(0.8)*	8.3(1.2)	94	17.3(2.1)	16.8(2.3)	17.6(2.3)	18.2(2.8)
36	8.1(0.9)*	7.7(1.0)*	7.8(0.9)*	8.7(1.3)	96	17.7(2.2)	17.1(2.4)	17.9(2.4)	18.4(2.8)
38	8.4(1.0)*	7.9(1.0)*	8.1(0.9)*	9.1(1.3)	98	18.1(2.3)	17.4(2.5)	18.3(2.5)	18.7(2.9)
40	8.6(1.0)*	8.2(1.0)*	8.3(0.9)*	9.5(1.4)	100	18.5(2.5)	17.8(2.6)	18.7(2.7)	19.1(3.1)
42	8.8(1.0)*	8.4(1.0)*	8.6(1.0)*	9.9(1.5)	102	18.8(2.7)	18.2(2.6)	19.1(2.9)	19.3(3.2)
44	9.1(1.0)*	8.7(1.0)*	8.8(1.0)*	10.3(1.6)	104	19.3(2.9)	18.5(2.6)	19.4(3.1)	19.5(3.2)
46	9.3(1.0)*	8.9(1.0)*	9.0(1.0)*	10.6(1.5)	106				
48	9.6(1.0)*	9.1(1.0)*	9.3(1.1)*	10.9(1.5)	108	19.9(3.1)	19.3(2.7)	20.3(3.3)	20.0(3.3)
50	9.8(1.1)*	9.4(1.1)*	9.6(1.1)*	11.1(1.5)	110	20.2(3.1)	19.6(2.8)	20.7(3.4)	20.5(3.5)
52	10.1(1.1)*	9.6(1.1)*	9.9(1.1)*	11.4(1.5)	112	20.6(3.2)	20.0(3.0)	21.1(3.6)	20.9(3.6)
54	10.3(1.1)*	10.0(1.1)*	10.2(1.1)*	11.6(1.6)	114	21.0(3.3)	20.5(3.1)	21.5(3.8)	21.3(3.6)
56	10.5(1.1)*	10.3(1.2)*	10.5(1.2)*	11.8(1.6)	116				
58	10.9(1.1)*	10.7(1.3)*	10.9(1.2)*	12.1(1.6)	118				
60	11.1(1.1)*	11.1(1.4)*	11.3(1.4)*	12.4(1.6)	120	21.8(3.3)	21.4(3.0)	22.5(3.8)	22.1(3.7)

BU-CD: Differential distance between subjects in group with word “Beautiful–Ugly” and “Close–Distant”
CM-BU: Differential distance between subjects in group with word “Classical–Modern” and “Beautiful–Ugly”
CD-CM: Differential distance between subjects in group with word “Close–Distant” and “Classical–Modern”
N: Differential distance between subjects in group without evaluation word
*Significant difference between this group and group without evaluation word

(3) Differential Distance Generated in Each Time Period

In this study, the differential distance generated per unit time (2 s) is referred to as the differential velocity. A faster velocity indicates that there was rapid increase in the differential distance between two observers in those 2 s; a negative velocity indicates that the differential distance between the two subjects is narrowing, that is, the heat maps are becoming similar. Figure 4 shows the trend of the differential velocity in the form of box plots. The figure demonstrates that the differential velocity was higher in the first 2 s, following a rapid decrease.

The top and bottom of a single box plot represent the maximum and minimum values of the dataset; the length of the line represents the dispersion of the dataset. Therefore, the mean of the differential velocity in each time period and the variance representing the dispersion in each time period were calculated using cluster analysis.

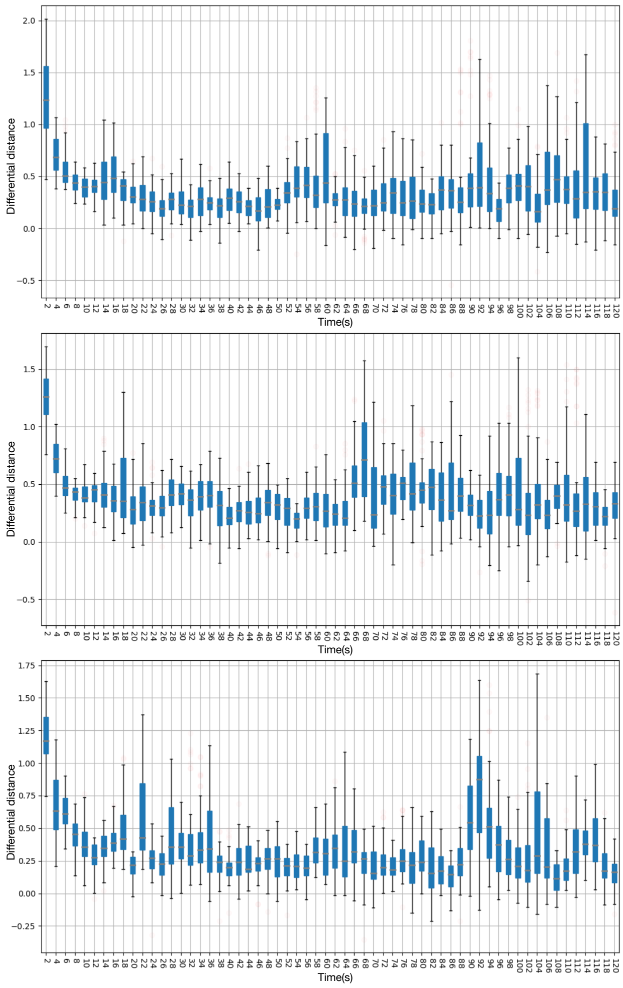


Figure 4: Differential velocity vs. time

Figure 5 shows the velocity and variance of each time period, and Table 5 shows the cluster under which each time period was categorized.

The 60 time periods were categorized into six clusters. The differential velocity of the time period in C1 and C2 was much higher than that of the other time periods. C3 had an average velocity and low variance. However, the velocity and variance under C4 were the least. Although the time period categorized under C5 and C6 had an average velocity, the variance was high.

The initial time periods of evaluation were categorized as C1 and C2. C3 often appeared at 6–18 s. In the “Close-Distant” group, it also appeared at 26–38 s and 70–84 s. The time period after 96 s was rarely categorized under C3. The time period categorized under C4 was often found in the “Beautiful-Ugly” group and “Classical-Modern” group. Overall in the three experimental groups, the time period from 38–56 s was generally categorized under C4. For the “Beautiful-Ugly” group, 66–88 s was often categorized under C4, and for the “Classical-Modern” group, 18–82 s was categorized under C4. C5 and C6 were not as common as C3 or C4. Most time periods categorized under C5 were 94–106 s, except for the “Beautiful-Ugly”; some time periods between 26–36 s were also categorized under C5. Most of the time periods categorized under C6 were after 86 s.

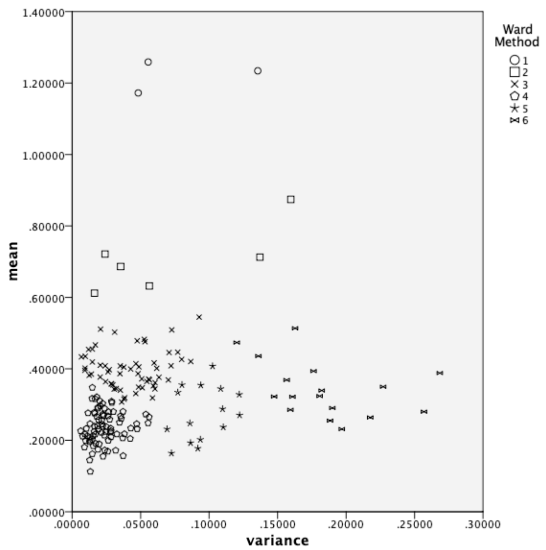


Figure 5: Variance vs. mean of each cluster

2.3 Discussion

(1) The Effect of Evaluation Words on Gaze Position During Evaluation

Because the heat map provides a more intuitive visualization of the gaze position, similarities and differences can be observed between different heat maps generated under the same evaluation word. The similarities of two heat maps can be quantified based on the differential distance between them.

It may be observed that the differential distance between subjects in the same group increased continuously. As mentioned above, a subject’s evaluation process is shaped by both universal thinking and individual thinking. When evaluation is based on the SD method, it can be inferred that different subjects are more inclined to observe and evaluate according to individual preferences, because there is no clear functional requirement for each part of the photograph, i.e., the influence of individual thinking on the subjects’ gaze is greater.

However, a comparison of the differential distance between the experimental groups and the control group revealed significant differences between them at particular time periods. The differential distance of the latter was often higher. The gaze position of the subjects with the same evaluation word would have similarities; therefore, it may be inferred that evaluation words affected gaze position. However, the influence was subject to the particular time period. The influence of different evaluation words varied according to the time period.

It can be conjectured that the evaluation words stimulated the subjects’ universal thinking. The similarity of the gaze position of different subjects might be attributable to a large extent to the evaluation word. While evaluating using the SD method, not only logic but also Kansei played a role in stimulating universal thinking. Therefore, it was established that the subjects’ Kansei was expressed through their gaze position.

Following the comparison of the differential distance of the subjects from the different experimental groups and the control group, it may be observed that even with different evaluation words, the differential distance still greater in control group. This is because different evaluation

Table 5: The cluster each time period categorized

TIME	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58	60
Beautiful–Ugly	1	2	2	3	3	4	4	3	3	4	3	4	4	5	3	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4
Close–Distant	1	2	3	3	3	3	3	3	5	4	3	4	4	3	3	3	3	3	3	4	4	4	4	3	4	4	4	4	4	4
Classical–Modern	1	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	6	6
TIME	62	64	66	68	70	72	74	76	78	80	82	84	86	88	90	92	94	96	98	100	102	104	106	108	110	112	114	116	118	120
Beautiful–Ugly	3	5	3	4	4	4	4	4	4	4	4	4	4	4	3	2	6	3	4	4	5	6	5	4	4	3	3	3	4	4
Close–Distant	4	4	3	2	5	3	3	3	3	3	3	3	5	3	4	4	5	3	5	6	6	6	4	3	6	6	5	3	4	3
Classical–Modern	4	4	4	4	4	4	3	4	4	4	4	3	3	6	6	6	6	4	3	3	3	5	6	6	3	6	6	3	3	5

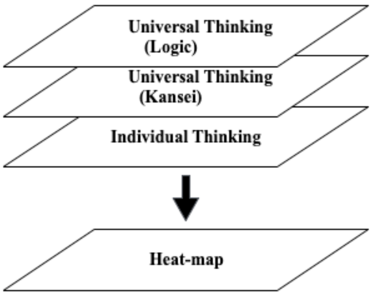


Figure 6: Three layers of heat map

words might have relationships with the same components or areas in a picture; the Kansei effect on the gaze position for different words also exhibited similarities.

Based on the analysis results above, the author deduces that each heat map generated for each subject is composed of three layers of heat map, as shown in Figure 6: the universal-thinking layer (logic), the universal-thinking layer (Kansei), and the individual-thinking layer. The universal-thinking layer (logic) heat map describes the universal area of all the subjects. It indicates the area that the subjects will naturally observe when executing the task of "evaluation", regardless of the evaluation words. The area described by this heat map may have a direct logical relationship with the task of "evaluation". The universal-thinking layer (Kansei) heat map describes the universal area of subjects assigned the same evaluation word. The area described by this heat map may have a direct logical relationship with the task of "Evaluate with this specific evaluation word." The individual-thinking layer heat map is mainly shaped by individual thinking, and there is no universal trend area among the subjects in this layer.

(2) The Effect of Universal Thinking and Individual Thinking Over Time

The comparison results based on the heat maps at different time periods demonstrate that the "power" of universal thinking varied over time. This can be inferred from the differential velocity. Because the differential distance steadily increased during the Kansei evaluation, when the differential velocity was low, the differential distance between the two heat maps varied gradually; the influence of universal thinking was inversely proportional to that of individual thinking. It is common knowledge that the direction and magnitude of velocity of an object in motion can be changed through the exertion of force. In terms of differential velocity, universal thinking and individual thinking are like two opposing forces exerted at the subject. The variance of the differential velocity can reveal whether the different subjects are under the same "force". The greater the variance, the more likely it is that the subjects are under the influence of different "forces".

From the cluster analysis result and the time period when each cluster appeared, the entire 120 s evaluation process can be divided into six stages. In Stage 1 (0–4 s), most of the time periods were categorized under C1 and C2, and the differential velocity was high. Stage 1 was the initial stage of observing the photographs; the different subjects' initial area of focus can be random; it can also be affected by individual thinking. Therefore, there was a higher differential velocity in this stage, because the subjects observed their preferred areas. In Stage 2 (4–18 s), of a total of 21 time periods in the three groups, 17 were categorized under C3. Although the differential velocity was still relatively high, it declined fast; therefore, the majority of the subjects were still dominated by individual thinking, and universal thinking began to exert a greater influence on the subjects. In Stage 3 (18–36 s), of the total 27 time periods in the three groups, 15 were categorized as C4, 8 were categorized as C3, and 4 were categorized as C5. In Stage 3, the differential velocity continued to decrease, and universal thinking began to exert a greater influence on the subjects' observation process, resulting in a large variance. In Stage 4 (36–56 s), of a total of 30 time periods in the three experimental groups, 27 were categorized as C4. In this stage, the velocity and the variance were at an overall low level. At this point, the universal thinking "force" could be said to be at its most potent during the evaluation. However, the velocity was still positive, also indicating that the differential distance was maintained, and that individual thinking was the dominant factor in the process of Kansei evaluation. In Stage 5 (56–86 s), of the total of 45 time periods for the three groups, 26 were categorized as C4, and 13 were categorized as C3. This stage was similar to Stage 3; it was the point at which the velocity increased from low to high. Some subjects began to be affected anew by individual thinking, resulting in greater variance. At this stage, it could be inferred that the subjects had completed the intelligence collection and rational analysis required for evaluation. In Stage 6 (86 s–the end of experiment), of the total of 51 time periods in the three groups, 15 were categorized as C3, 13 as C4, and 15 as C6. In this stage, the velocity increased, and the variance became high, indicating the randomness in the influence exerted by individual and universal thinking were scattered. It was speculated that at this stage, some subjects had completed the evaluation task, and might have lost the purpose of observation, while others might still have been carrying out the evaluation task and were slightly influenced by universal thinking.

Generally, the entire evaluation process was greatly affected by individual thinking because the differential distance increased continuously. However, from Stage 2, universal thinking began exerting its influence, which peaked in Stage 4 and subsequently decreased. During the last stage, it was impossible to determine the obvious trend as the time required for the individual evaluation task varied, which led to great differences among the subjects.

3. EXAMPLE OF ANALYZING FACADE THROUGH HEAT MAPPING

After elaborating on the evaluation process and structure of the heat map, the Kansei expressed by the gaze position of the subjects during the evaluation can be further discussed by extracting the heat maps from each layer and comparing them with the photograph used in the actual evaluation.

3.1 Method

Following the experiment, it was found that during the evaluation process, although the observation results of the subjects were mainly affected by individual thinking and some random factors, the effect of universal thinking on the subjects peaked in Stage 4; therefore, the heat map at the end of Stage 4 was the most suitable for analyzing the subjects' universal thinking.

Furthermore, the analysis results above revealed that all the subjects had a universal observation area during evaluation. These universal areas were the universal-thinking layers. To analyze the universal-thinking layer heat map, eight heat maps numbered 1–8 were extracted from the heat maps of all the subjects at the end of Stage 4; their relationship is shown in Figure 7, and the description of each heat map is interpreted in Table 6.

Two algorithms executed using Python were deployed for the extraction of the heat map. The methods were the similarity extraction and difference extraction. The former could extract the universal area and dynamic values from all the heat maps to be processed and convert them into one heat map. The latter could extract the different areas and dynamic values from two heat maps and convert them into one heat map. The calculation procedures of both processes are detailed in Figures 8 and 9. The extraction process of the heat maps numbered 1–8 is detailed in Table 7.

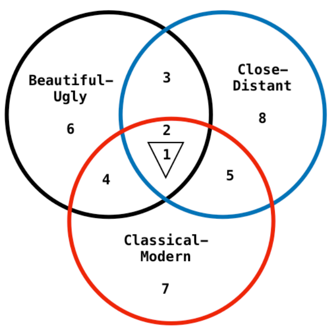


Figure 7: Relation of each heat map

Table 6: Description and inferred meaning of each heat map

No.	Description of heat map	Inferred Meaning	Layer
1	Areas observed by all subjects	Area affected by the task "evaluate" regardless of the evaluation words .	Universal-thinking layer (logic)
2	Areas observed by all subjects who had a evaluation word (excluding the area in 1)	Area affected by universal thinking stirred by the three evaluation words, "Beautiful–Ugly", "Close–Distant", and "Classical–Modern"	Universal-thinking layer (Kansei)
3	Areas observed by all subjects who had the evaluation word "Beautiful–Ugly" and "Close–Distant" (excluding the area in 1,2)	Area affected by universal thinking with the evaluation words "Beautiful–Ugly" and "Close–Distant"	
4	Areas observed by all subjects who had the evaluation word "Beautiful–Ugly" and "Classic–Modern" (excluding the area in 1,2)	Area affected by universal thinking of both "Beautiful–Ugly" and "Classical–Modern"	
5	Areas observed by all subjects who had the evaluation word "Close–Distant" and "Classic–Modern" (excluding the area in 1,2)	Area affected by universal thinking of both evaluation words "Close–Distant" and "Classical–Modern".	
6	Areas observed by all subjects who had the evaluation word "Beautiful–Ugly" (excluding the area in 1,2,3,4)	Area solely affected by universal thinking of evaluation word "Beautiful–Ugly"	
7	Areas observed by all subjects who had the evaluation word "Classic–Modern" (excluding the area in 1,2,4,5)	Area solely affected by universal thinking of evaluation word "Classical–Modern"	
8	Areas observed by all subjects who had the evaluation word "Close–Distant" (excluding the area in 1,2,3,5)	Area solely affected by universal thinking of evaluation word "Close–Distant"	

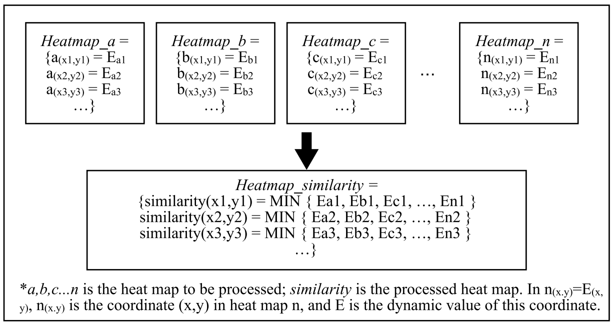


Figure 8: Similarity extraction

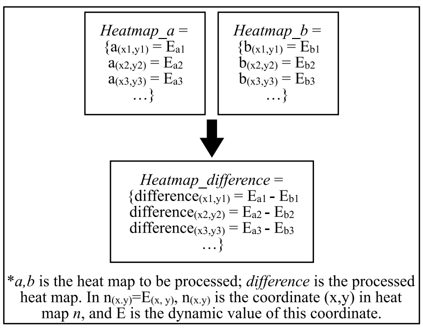


Figure 9: Difference extraction

Table 7: Extraction process

Step	Heat map Used	Method	Results
1	All subjects	Similarity	Heat map 1
2	All subjects who had evaluation words	Similarity	Superposed Heat map by Heat maps 1 and 2
3	[Superposed Heat map by Heat maps 1&2] and [Heat map 1]	Difference	Heat map 2
4	All [Beautiful–Ugly] group subjects	Similarity	Superposed Heat map by Heat maps 1&2&3&4&6
5	All [Classical–Modern] subjects	Similarity	Superposed Heat map by Heat maps 1&2&4&5&7
6	All [Close–Distant] subjects	Similarity	Superposed Heat map by Heat maps 1&2&3&5&8
7	All [Beautiful–Ugly] & [Classical–Modern] subjects	Similarity	Superposed Heat map by Heat maps 1&2&4
8	All [Beautiful–Ugly] & [Close–Distant] subjects	Similarity	Superposed Heat map by Heat maps 1&2&3
9	All [Classical–Modern] & [Close–Distant] subjects	Similarity	Superposed Heat map by Heat maps 1&2&5
10	[Superposed Heat map by Heat map 1&2&3] and [Superposed Heat map by Heat map 1&2]	Difference	Heat map 3
11	[Superposed Heat map by Heat map 1&2&4] and [Superposed Heat map by Heat map 1&2]	Difference	Heat map 4
12	[Superposed Heat map by Heat map 1&2&5] and [Superposed Heat map by Heat map 1&2]	Difference	Heat map 5
13	[Superposed Heat map by Heat map 1&2&3&4&6] and [Superposed Heat map by Heat map 1&2&4]	Difference	Superposed Heat map by Heat maps 3&6
14	[Superposed Heat map by Heat map 1&2&4&5&7] and [Superposed Heat map by Heat map 1&2&5]	Difference	Superposed Heat map by Heat maps 4&7
15	[Superposed Heat map by Heat map 1&2&3&5&8] and [Superposed Heat map by Heat map 1&2&3]	Difference	Superposed Heat map by Heat maps 5&8
16	[Superposed Heat map by Heat map 3&6] and [Heat map 3]	Difference	Heat map 6
17	[Superposed Heat map by Heat map 4&7] and [Heat map 4]	Difference	Heat map 7
18	[Superposed Heat map by Heat map 5&8] and [Heat map 5]	Difference	Heat map 8

3.2 Result

A total of eight heat maps were extracted using the above-mentioned methods; the extracted heat maps are shown in Figure 10. These heat maps can visually show the gaze position under different evaluation words; the heat maps would be superimposed due to the analysis purpose. Subsequently, this study will illustrate the process detailed so far by analyzing the observation area of the different evaluation words used in the experiment.

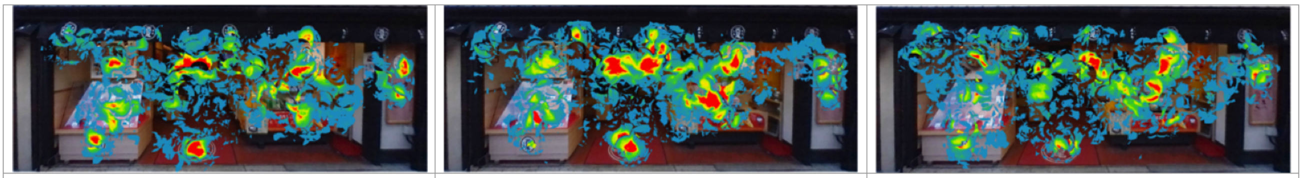
3.3 Discussion

(1) Universal Thinking (Logic) During Evaluation

Based on the similarity extraction, it was possible to utilize the heat maps of all the subjects for the analysis of the universal thinking (logic) of evaluation. Heat map 1 is the common observation area of all the subjects. Regardless of the presence or absence of evaluation words, the area marked by Heat map 1 was observed. It can be said that these areas were the universal logic when the subjects executed the task of "Evaluate the store facade". The heat map revealed that the subjects' gazes were more evenly distributed in the central area of the entire store, whereas they were not distributed in the surrounding walls, floors, and related areas. In the area where the gaze was distributed, the right-side commodity had the highest dynamic value.

The dynamic values of the light box, television on the inside wall of the store, menu posted on the left partition, food model cabinet on the left, carpet at the entrance, and board on the outside were all slightly high, which is consistent with the previous speculation that logical

Figure 10: Extracted heat map nos. 1–8



Heat-map superposed by Heat-map No.2&3&4&6. Used to analyze the word "Beautiful-Ugly".

Heat-map superposed by Heat-map No.2&4&5&7. Used to analyze the word "Classical-Modern".

Heat-map superposed by Heat-map No.2&3&5&8. Used to analyze the word "Close-Distant".

Figure 11: Superposed heat map

layers mainly correspond to the state of scanning the entire store. Because the area to which the bulk of observation time was devoted was largely in the right commodity area, it can be deduced that the commodity was always of greatest interest to the subjects during the facade evaluation.

(2) Universal Thinking (Kansei) of Each Word

When analyzing the effect of the evaluation words on gaze, the related heat maps of all the universal thinking (Kansei) layers were superimposed and utilized. The superimposed ones are shown in Figure 11.

Heat maps 2, 3, 4, and 6 are related to “Beautiful–Ugly”. On this basis, the areas with high dynamic values were mainly the carpet at the entrance of the store, logo downside the food model cabinet, poster above, second logo on the right side of the curtain, board outside the store, and space inside the store. It is apparent that the subjects were inclined to look at the objects near the store entrance and objects with logos when the evaluation word was “Beautiful–Ugly”.

Heat maps 2, 4, 5, and 7 are related to “Classical–Modern”. On this basis, the areas with high dynamic values were concentrated in the interior space of the store and on objects such as traditional plaques, television, Japanese-style partitions, curtains, and light boxes inside the store. In addition, the commodity areas and carpets at the entrance also had high dynamic values. The combination of classical plaques and modern objects such as the television in the interior space may have attracted the attention of the “Classical–Modern” group.

Heat maps 2, 3, 5, and 8 are related to “Close–Distant”. The red-colored area in the heat map is smaller than that of the other two words. It can be inferred that the similarity in the gaze of the subjects assigned this evaluation word was relatively low. The objects with higher dynamic values were the television, light box, posters above commodity and food model platforms, and board outside the store. These objects include commodities and store information and were highly interactive.

4. CONCLUSION

The gaze of subjects evaluating the photographs of store facades using the SD method was analyzed, and the observation process was discussed using chronological heat mapping. The heat map was decomposed into three layers: the universal-thinking layer (logic), the universal-thinking layer (Kansei), and the individual-thinking layer. The similarities of gaze position associated with each word were extracted by two algorithms based on the differences and similarities of the heat maps.

Although the relationship between the heat map and users’ Kansei could not be clearly measured, the area of focus during the evaluation process could be extracted from the heat map. Additionally, it was possible to speculate the areas that were most likely to hold the subjects’ attention, a possibility with great potential for the evaluation process in various fields in the future. In future studies, the author will undertake research to clarify the relationship between users’ gaze and Kansei. Furthermore, the information inherent in the gaze goes beyond the gaze position; the author intends to excavate this information and discuss it in future studies.

REFERENCES

1. Ichihara, S.; The perspective of the research for the semantic differential and the problems to be solved, The Japanese Journal of Ergonomics, 45(5), pp.263-269, 2009. (in Japanese)
2. Yoshioka, K.; Impression evaluation and eye movement related to the characteristic expression as elements in abstract paintings: Mondrian, Malewitsch and Rothko, Kansei Engineering International Journal, 10(1), pp.81-89, 2010.
3. Köhler, M., Falk, B., and Schmitt, R.; Applying eye-tracking in Kansei engineering method for design evaluations in product development, International Journal of Affective Engineering, 14(3), pp.241-251, 2015.

4. Siirtola, H., and Räihä, K.-J.; Using gaze data in evaluating interactive visualizations, In: Ebert, A., Dix, A., Gershon, N. D., and Pohl, M. (eds), Human Aspects of Visualization, HCIV 2009, Lecture Notes in Computer Science, 6431, pp.127-141, 2009.
5. Jacob, R., and Karn, K.; Eye tracking in human-computer interaction and usability research: Ready to deliver the promises, In: Hyöna, J., Radach, R., and Deubel, H. (eds.); The Mind's Eye: Cognitive and applied aspects of eye movement research, Elsevier, Amsterdam, pp.573-604, 2003.
6. Steichen, B., Carenini, G., and Conati, C.; User-adaptive information visualization - Using eye gaze data to infer visualization tasks and user cognitive abilities, IUI '13 Proceedings of the 2013 International Conference on Intelligent User Interfaces, pp.317-328, 2013.
7. Shimizu, Y.; Kansei kougaku no houhou (The methodology of Kansei engineering), Journal of the Society of Biomechanisms, 40(1), pp.5-11, 2016. (in Japanese)

Chenyang LIU (Member)
Chenyang Liu is a Ph.D student of Graduate School of Integrated Frontier Sciences, Kyushu University, Japan. His research interests include Kansei science and the methods of capture and quantify users' Kansei.

Haruka SOGABE (Member)
Haruka Sogabe is an Associate Professor of Faculty of Design, Kyushu University, Japan. Her interests are Kansei Design, Public Design and Industrial Design.